

Supplemental Protocol

Genechip Data Analysis

Absent/Present calls were generated from scanned arrays using Affymetrix GCOS 1.4 software and all subsequent analyses were conducted using dChip software as of April 2007 (<http://www.dchip.org>, Wong Lab, Harvard). For each sample type (sperm cells, pollen and seedling), a microarray experiment was performed with three biological replicates. We used a samplewise normalization to the median probe cell intensity (CEL) of all 9 nine arrays, and for each sample, the median CEL intensity of one replicate was scaled to the median CEL intensity of all arrays and defined as baseline. The remaining replicates of each sample were normalized to the baseline applying an Invariant Set Normalization Method (Li and Wong, 2001). Model-based gene expression was obtained from normalized CEL intensities based on a Perfect Match-only model (Li and Hung Wong, 2001). Only genes called Present in at least two replicates of each sample type were considered Present. For downstream analysis we used only 16,385 genes, which were filtered according to their consistent expression levels within replicates (called Present within a variation of $0 < \text{median} (\text{SD}/\text{mean}) < 0.5$) and if called Present at least in one of the 9 arrays. Annotations for all the genes represented on the Arabidopsis ATH1 genome array were obtained from the NetAffx database (www.affymetrix.com) as of June 2007 and manually updated according to TAIR7 as of May 2007.

LITERATURE CITED

- Li C, Hung Wong W** (2001) Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biol* **2**: R32
- Li C, Wong WH** (2001) Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A* **98**: 31-36